

# 統計学習における学習者のPPDACの実践

PPDAC experiences of students in statistics learning

中 村 宗 敬\*

NAKAMURA Munetaka

**要約**：数学の一分野の科目として見ると、統計学の学習においても、講義を出発点に、問題演習をへて評価として試験または演習問題中心のレポートを課するという理論重視の方向性が考えられる。一方で、この方法は学習者の関心・意欲をひきおこし、かつ持続させるのが困難である。この観点から筆者の授業実践として最後の仕上げに、PPDACサイクルを意識し自らそれを実践した上でのレポートを課した。その内容と成果について報告するが、履修者は概ね自らの学習、特にレポート作成時の問題発見、データ探索過程、その解釈・吟味という未経験の過程を有意義に受け止めていた。

**キーワード**：PPDAC, 統計思考, ポアソン分布, サッカーの得点分布

## I はじめに

本論は筆者の所属する教員養成学部・数学教育専門課程での統計学の授業実践に関わる報告である。授業名称は「数理統計学」であり、かつては

教科書に沿った講義→問題演習→試験またはレポート

というコースを行っていた。「数理統計学」といかにも理論重視の名称もその名残であり、他の数学の科目と同じく、理論を重視した伝統的な方法に則っていたのである。流布している学部学生用の教科書もそれに沿って書かれている（反面、応用系数学分野の教科書によく見られることだが、数学的厳密性は中途半端である。例えば、最小2乗法で回帰直線を求める際には、偏微分=0の方程式を解くのみで、最小性に関する論証を欠いている）。扱う問題は、数値データとしては初学者に扱いやすい、手計算、もしくは単純機能のみを持つ電卓でも計算可能である程度の人工的なものがほとんどである。筆者のかつての授業では、総仕上げとしての試験または演習問題を中心にしたレポートも、それらに類したものであった。

しかし、他の数学科目と異なり、現実に出会った問題をどう解決するか、という本来の持つ統計的思考力養成の側面がこのような学習コースでは活かされず、かつ、学習者の意欲を削ぐことになりかねないと考える。実際に、授業アンケートあるいは学生からの直接の反応等からもそれを察することができた。教科書による基礎学習の重要性を認めるにしても、そのみでなく統計学習独自の良さを引き立たせ、学習者にとって喫緊の必要性として教育実習で現実的なデータを扱う授業を行う際に役立たせる（当該の「数理統計学」は2年次の科目で教育実習の前年に開講されている）ような枠組みを設けるべく模索した。

結果的には上述の従来の統計学習で足りない部分を補う手段として、統計的思考力養成の最終目標としてレポートを課した。ここでいうレポートは、単なる教科書の練習問題・章末問題に類するものの解答のそれではなく、次項目を各自で決めさせるという学習者の興味・関心と自主性を重視

\*教育科学域 人間科学系

したものである。

- (1) 興味を持った現実事象の問題を自分で探し、統計問題として自主的に問題設定する
- (2) 問題に合ったデータ収集計画をたてる
- (3) データを自力で調べる
- (4) 学習事項のどのような統計手法が問題解決に有効かを試行錯誤する
- (5) 統計としての結論を出し、現実と対比する

これは、近年において統計調査実践において非常に重要視されている基本的概念の PPDAC ([2], [5])

**Problem→Plan→Data→Analysis→Conclusion (→Problem→…)**

のサイクルを意識させ、最終目的のレポートにおける統計調査もこの趣旨にしたがって行うことを心がけるよう伝達した。

レポートに望むに際しては、サッカーに興味があるとしてそこを出発点とした上記 PPDAC サイクルの仮想例を示した。これは後に述べる。レポートに書かれた感想として、データ収集の大変さを多数の者が訴えていた（これは、情報機器使用が学習の組み込まれていなかったことも一因である）。一方で、疑問に感じていたことを自主的に問題設定を行い、限られた知識の中でも自力解決に至ることができて楽しかったし、有意義であったという記述も多かった。

## II 授業の全体設計

平成28年（2016年）度に行った「数理統計学」の授業実践について述べる。履修者は17名で、ほとんどの者が前期にこの授業の理論的基礎となる「確率論」を受講している。この程度の少数であることは授業者としてはやりやすい環境で、ある程度学習者への直接関与も可能である。確率を含めるとはいえ通年で統計関係の内容に取り組むことができるので、時間的な余裕も大きい。

さて、PPDACにより学習者主体でレポート作成に臨むといっても、その基礎になる統計知識が不可欠である。そこで、15回を二つの部分に分け、

第一部：教科書に準拠した基礎的事項の学習

第二部：レポートに関する注意の伝達と作成例の提示

の構成にした。二分といっても、講義15回のうち、10回程度の時間を第一部に費やしている。

肝心の第二部であるが、レポート作成に関してはごく簡単なあらましを授業開始時に述べ、第一部学習の後に変えてもよいから、とにかく問題探索を早く意識するように促した。

第一部では教科書を網羅的に授業で扱うことはしていない。扱った内容は

### 1. 記述統計

- a. 高校時の「データの分析」
- b. 回帰直線と相関係数の関係

### 2. 推測統計

- a. 推定  
正規分布、 $t$ 分布を使った母平均の推定、正規近似による比率の推定
- b. 検定  
上記推定に対応する検定

の各項目である。きわめて基礎的な知識のみにとどめている。統計的処理の基礎になる概念や方法の伝達に努め、教示中心にならないようワークシートを用いたアクティブラーニング的要素を随所に入れていく。

標準的にどの教科書にも載っている項目のうち、

- ・  $F$  分布を用いた分散の推定, 検定
- ・ 2 種類のデータの平均の差の推定, 等平均の検定
- ・  $\chi^2$  分布を用いた適合度検定

は第一部では扱っていない。このうち、三番目の「 $\chi^2$  分布を用いた適合度検定」は、次節において述べるように、第二部の内容提示の際に用いているのでそこで説明している。とにかく無理に全体場で学習範囲を広げず、問題設定や分析の段階で必要になったら、それに応じた方法を教科書等で学ばよという考えである。

第二部で登場するレポートに関しては、どのような形態のものにするのかを授業設計段階で試行錯誤している。まず、問題点として考えたのは、完全に問題発見・設定から解決に至るまですべて学習者側に任せてよいかということであった。こちらが出発点となる共通の問題設定を行い、現実事象から得られるデータを提示して、その後における PPDAC サイクルの複数回の展開は各自に任せるという方法も一案としてあった。しかし、最終的に問題発見と設定から解決に至り、そこから再度新たな問題発見をし、サイクル的に調査を行うことの重要性を優先した。それが問題への関心を持続させると考えたからである。このような学習経験はほとんどの者にとって初めてであり、筆者としては、この機会を探求的・研究的態度を養成する端緒となることを意図している。

### III レポート作成例の提示

前節の第二部で提示したのは、実際のレポートでない。あくまでも仮想的の統計的思考・調査の例として筆者が作った教材である。以下それを記すが、まず、仮想のレポート作成者がサッカーに興味を持っている、非常に好きであると仮定して、次のような疑問、すなわち PPDAC の P から始める。以下では、記述主体はこの仮想レポート作成者と想定する。

**Problem** この作成者が友人から「サッカーの試合、特にプロにおいては得点が少なく面白味にかけると言われる。そこで、「得点が少ない」というのは本当なのか、と疑問を持ち、調べてみることにした。

**Plan** 日ごろテレビで視聴してる最も有名なサッカーリーグであるスペイン 1 部の 1 試合の得点状況を調べてみることにした。

**Data** ネットで当該サイトを検索し、データを集める。

**Analysis** 1 試合の得点数ごとの延べチーム数を度数分布表にまとめる。次の表を得た。

表 3. 1 サッカーのスペイン 1 部リーグの得点分布

得点数	0	1	2	3	4	5	6	7	8	9	10
度数	214	259	166	72	25	15	6	1	1	0	1

リーグ所属チームは 20 チームあるので、ホームアンドアウエーを考えると調査対象の延べチーム数は 760 チームとなる ( $2 \times 20 \times 19 = 760$ )。そのうち、ある試合の結果が 1-0 ならば、得点数 1 の度数に 1, 得点数 0 の度数に 1 を数える。これを全 380 ( $= 20 \times 19$ ) 試合についてまとめた結果が上の表である。

この度数分布表から代表値として平均（値）を計算すると、1.372が得られる。

**Conclusion** 度数分布表で得点数 0, 1 の階級でほぼ 6 割を占め、全体として低得点の傾向がみられる。平均も 1.372, すなわちチームごとに見ると、1 試合 90 分以上の中でこれくらいしか得点する場面が見られないことがわかる。テレビ視聴者も含めて観客はひいきチームの側の得点を期待し、実際それが成功すると喜び興奮するのであるから、これでは確かに「面白くない」という見方ももつともであろう。

これで PPDAC の 1 サイクルが完結した。しかし、これでは大学での統計学習の成果としては素朴すぎ（中学生でもこの程度は分析可能である）、2 節で述べた基礎学習内容の成果が生かされていないので、次のような新たな疑問を喚起する。

**Problem** 表 3.1 のようなサッカーの得点分布はポアソン分布に近いということを確率論で学んだ。これはスペインリーグでもあてはまるのだろうか。

**Plan** 前期の「確率論」では 2014 年ワールドカップの得点分布の相対度数をヒストグラムにして、ポアソン分布の推定値と比較参照すると、グラフは非常に似通ったものになった（実際、この年度の前期の「確率論」でこの内容を扱っている）。

**Data** データはすでに得ている表 3.1 を利用する。

**Analysis** 表 3.1 から相対度数を計算し、平均が 1.372 のポアソン分布と比較したのが次の表 3.2 である。

表 3.2 スペイン1部リーグの得点ごとのチーム相対度数と平均 1.372 のポアソン分布値の比較

得点数	0	1	2	3	4	5	6	7	8	9	10
相対度数	0.282	0.341	0.218	0.095	0.033	0.020	0.008	0.001	0.001	0.000	0.001
ポアソン	0.254	0.348	0.239	0.109	0.037	0.010	0.002	0.000	0.000	0.000	0.000

延べ総チーム数が 760 であるから、相対度数は、度数 / 760、ポアソンの欄の値はポアソン分布の定義式に従って、 $e^{-\lambda} \lambda^x / x!$ ,  $x=0,1,2,\dots$  ( $\lambda=1.372$ ) と表される。これを数値計算し「ポアソン」欄に記した。

両欄の数値が近いことが見てとれるが、さらに比較しやすくするためにグラフを作成する（次ページの図 3.2）。これを見ると、確かにヒストグラムと折れ線は非常に似た傾向を示している。授業の第 1 部では触れなかったが、教科書に適合度検定というものがある、データと理論的に期待される分布が同じものであるかを統計的に検定する方法が書かれている。これを適用してみよう。

表 3.3 適合度検定に際しての表 3.1 のカテゴリー合併化とポアソン分布による期待度数

得点数	0	1	2	3	4	5以上
度数	214	259	166	72	25	24
期待度数	192.664	264.407	181.432	82.997	28.476	10.025

通常の手順に従ってあまり期待度数 (= 760 × ポアソン) が小さくならないようにいくつかの得点数のカテゴリーを合併する。具体的には、5 点以上をまとめて、表 3.3 のようにする。これより以下の  $\chi^2$  値を計算し、これと有意水準 0.05 の棄却域下限  $\chi^2_{4}(0.05)$  とを比較する (自由度はパラメーター  $\lambda$  の推定のため、6-1-1=4 となる) :

$$\chi^2 = \frac{(214-192.664)^2}{192.664} + \frac{(259-264.407)^2}{264.407} + \frac{(166-181.432)^2}{181.432} + \frac{(72-82.997)^2}{82.997} + \frac{(25-28.476)^2}{28.476} + \frac{(24-10.25)^2}{10.25} = 25.150,$$

$$\chi^2_{4}(0.05) = 9.488.$$

これにより、

「帰無仮説：スペイン一部リーグの得点分布はポアソン分布に従う」が有意水準 0.05 で棄却される。

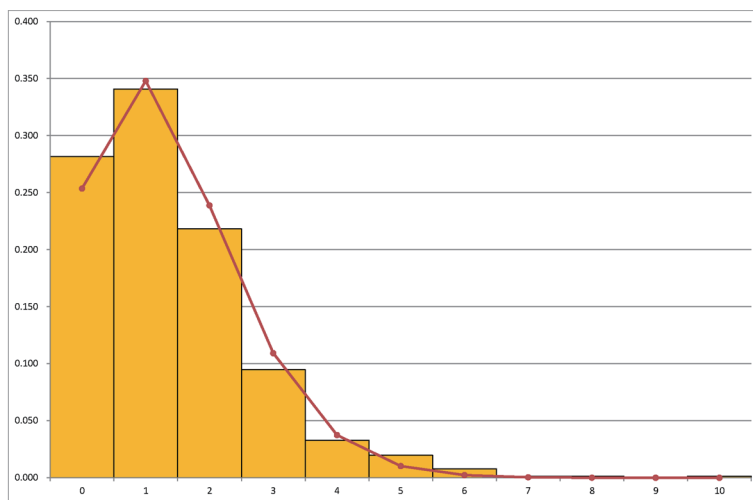


図 3.1 表 3.2 のグラフ化

ヒストグラムは表 3.1 の「相対度数」の値、折れ線は同表の「ポアソン」の値を表す。折れ線中の整数に対応する丸点以外では意味はないから折れ線化は適当とはいえないが、見やすくするために上のようになっている (これは離散分布のグラフ化の際、しばしば使われる。次節のグラフでも同様である)。

以上の提示をしたのだが、2 番目のサイクルの Conclusion はあえて説明せず、これがスペイン一部リーグのどのような状況を示唆しているかを各自考えるよう促した。ポアソン分布にしたがっていない要因としては、度数分布表 3.1 が高得点側によりすぎていることがある (0 点の度数もやや多い)。これはリーグのチーム間の強弱の差が大きいことに依っているのだろう。詳細を見ると特に、強いチーム (2 チームが抜きん出て強いことは有名であろう) がホームで下位チームと対戦した場合、大量に得点をするのがしばしばある。

なお、ポアソン分布にしたがわないもう一つの簡単な判別方法は、表 3.1 の分散を計算することである。ポアソン分布は平均と分散が一致しているから、もしポアソン分布にしたがっていればデータから得られる分散も平均 1.372 に近いはずである。もう少し詳しく見ると、分散の検定を行うことになるが、ここではそれは省略する。授業でも簡単に言及したのみである。



## IV 学習者によるレポート例

一つのレポート例を述べる。これは前節で述べた提示例と同じく、サッカーの得点分布に関する考察である。提示例のまねをしたものにはすぎないように見えるが、興味深いデータ解析も加えているので紹介することとした。

この学習者をAとする（イニシャルでもない単なる記号化した仮名である）。Aは山梨から遠く離れた土地の出身で、その地元チームが所属するプロサッカーリーグJ3（所属チーム数は16、したがって延べ総チーム数は $2 \times 16 \times 15 = 480$ である）について、前節と同様にチーム1試合あたりの得点分布を調べた。内容的には同じなので表3.2に対応する次表4.2から出発する。実際のレポートは表3.1の度数分布表から始めていた。

表4.1 サッカーJ3の得点ごとのチーム相対度数と平均1.233のポアソン分布値の比較

得点数	0	1	2	3	4	5	6	7
相対度数	0.288	0.373	0.215	0.079	0.038	0.006	0.002	0.000
ポアソン	0.291	0.359	0.222	0.091	0.028	0.007	0.001	0.000

相対度数は度数/460により得られる（度数分布表は省略している）。平均は1.233であり、これを平均とするポアソン分布の値を「ポアソン」欄に記している。

また、これも同じく相対度数とポアソン分布をそれぞれヒストグラムと折れ線とで比較したのが、次の図4.1である。

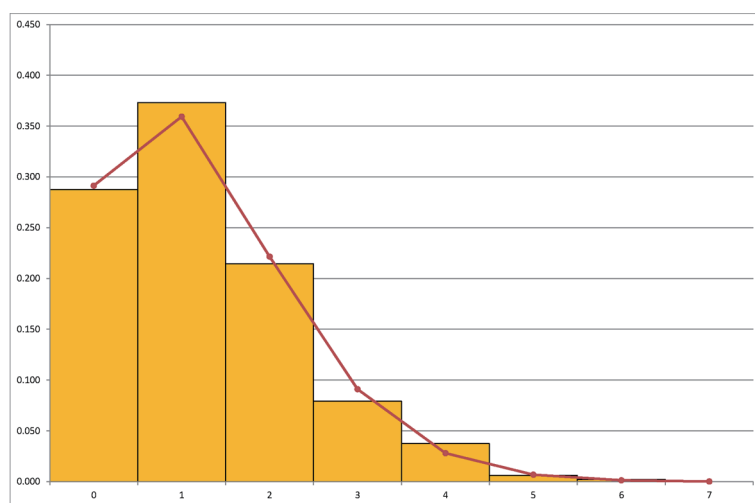


図4.1 表3.1のグラフ化

ヒストグラムが表3.1の「相対度数」、折れ線が「ポアソン」を表している。

この視覚化の後、さらにAは適合度検定を行っている。詳細は省くが、今度はJ3の得点分布はポアソン分布にしたがっているという帰無仮説は否定されない。グラフでも見てもスペインの高得点側に長い裾野はない。

ここまでは結果は逆になっているが、内容としては提示例とほとんど変わらない。その後Aは次のようなことをあらためて調べている。Aの地元チームであるZはJ3でも下位に低迷しているチームであるが、Zの得点状況は全体の中でどうなのだろうか調べたのである。それをまとめたのが表4.2である。

表 4. 2 チーム Z の得点ごと相対度数と平均 1. 233 のポアソン分布値の比較

得点数	0	1	2	3	4	5	6	7
相対度数	0. 400	0. 333	0. 200	0. 033	0. 000	0. 033	0. 000	0. 000
ポアソン	0. 291	0. 359	0. 222	0. 091	0. 028	0. 007	0. 001	0. 000

相対度数はその得点の試合数/30, 比較のため既出のポアソン分布の値も併記した。

さらにこれを前と同様にグラフ化したものが次の図 4. 2 である。Z の平均は 1. 00 である。

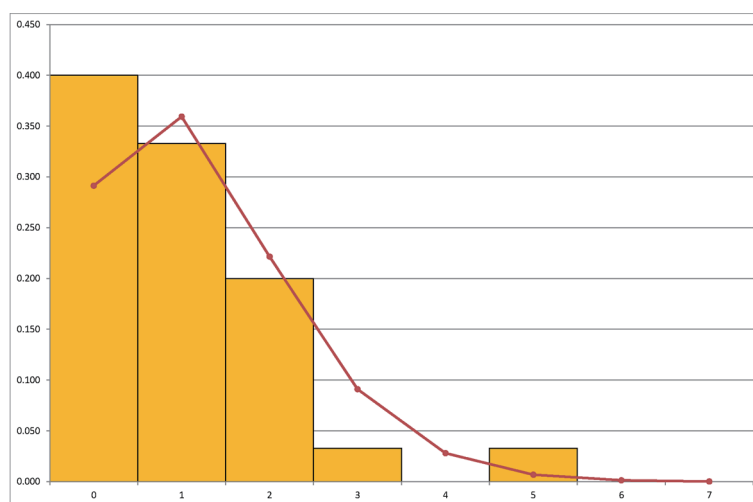


図 4. 2

ヒストグラムが表 3. 1 の「相対度数」, 折れ線が「ポアソン」を表している。

データ数が 30 と少ないのでかなり崩れた形ではあるが, 全体のポアソン分布から低得点側にずれていることが読みとれる。A はこれから, Z は得点が少ない。最終的な結果順位から予想されたが, それがはっきりしたと結論を下している。

## IV 考察および今後の課題

なぜ前節で A の作成レポートを紹介したのかをここでさらに説明する。ポアソン分布は表の確率が小さく試行回数が大きい二項分布の近似として得られる。これは前期の「確率論」でも論じている。したがってサッカーの得点分布がポアソン分布にしたがう, あるいはそれに近いということの意味は, サッカーの得点状況と稀にしか表が出ないコイン投げとほぼ同一視できるということにはほかならない。しかし, コイン投げというと, 同じコインを投げ続けるのが普通の捉え方であり, それから類推してリーグ内でどのチームも得点できる確率はすべて同じであるという結論は納得しがたいであろう。稀にしか出ないコイン…これらは稀の中でも, 起こりやすい, 起こりにくいとの差がある…のいくつかを投げ続けていることと同一視できるというのが事の真相であり, ポアソン分布に従う得点状況でも強いチーム, 弱いチームの得点状況がならされた結果であることを理解する契機となりうるという意味で, A のレポートを筆者は高く評価している。これはかなり専門的になるので A はそのことをはっきりとは捉えきれてはいないが, そこは学部 2 年次であることを考慮して今後期待したいところである。

なお, 数学的には, 上に述べたことは異なるパラメーターのポアソン分布の和の分布が, パラメーターの和を新たなパラメーターとするポアソン分布に近づくということに対応している。これ

については [1], [3], [4] を参照のこと。

今後の課題を三点上げておく。これは学生側というよりは授業者としても筆者の側の課題である。まず、一点目は情報機器の使用を学習の中に組み込む必要があるということである。学生たちは情報機器を用いた教育をこれまでも受けてきているが、まったくいかしきれていない。計算を電卓で行ってそのすべての数式処理過程を延々と書き上げたり、グラフも手描きのものが少なくなかった。極端なものになると方眼紙さえ使わない。データ収集時の大変さは別にしても、その処理に関する労力は大幅に低減されるはずである。

二点目はレポートの書き方である。ちいさな論文の形にせよ、と要求はしてその簡単な説明もしたのであるが、残念ながらそれが伝わっていない。形式のみではなく文章表現そのものもたどたどしいものが散見される。この授業にレポートの書き方を学習内容に組み込むのは本意ではないが、履修者に共通するそのような場がないとしたら、考慮に入れなければならない。

三点目は授業設計の一層の改善である。基礎と応用を二分したことで、その連携を滑らかにする必要がある。この年度の授業では唐突に第二部に入っていったという感があるし、学生もそのように言っていた。基礎を学ぶ第一部の段階から教科書以外の生のデータを提供して現実問題を意識させることも必要なのではないかと考えている。

第1節でも述べたが、学生たちは自己評価ながら、統計的思考力、データを見きわめる力を高めることができたという学生が多かった。この場では紹介できなかったが筆者の興味をひいたいくつかのレポート題目を記しておく。

- ・お年玉年賀はがきの番号の一様性について
- ・生菓子の全商品の売上との相関
- ・J3全体と覇員チームの得失点分布の比較
- ・コンビニ点数と人口の相関

慣れない作業に苦勞しながらレポート作成を進め、結果として興味深いものを多く提供してくれた彼らに感謝したい。

#### 参考文献

- [1] Durrett, *Probability: Theory and Examples*, Wardswaorth&Brookes/Cole, Belmont,CA, 1991.
- [2] Wild, C.J., Pfannkuch,M., Statistical Thinking in Empirical Enquiry, *Intrnational Statistical Review* (1999). 67, 3, 223-265.
- [3] フェラー, W., 確率論とその応用 下, 第XII章, 紀伊國屋書店, 1961.
- [4] ブロム, G., ホルスト, L., サンデル, D., 確率論へようこそ, 第8章, シュプリンガー・フェアラーク東京, 2005.
- [5] 渡辺美智子, 知識基盤社会における統計教育の新しい枠組み～科学的探究・問題解決・意思決定に至る統計思考力～, 日本統計学会誌, 42, 2, 253-271.